[6] L. G. Mason and T. Gavin, "An application of optimal control theory to modernization problems in telephone networks," presented at Optimiz. Days Conf., Montreal, P. Q., Canada, 1978.

[7] T. Gavin and L. G. Mason, "Some models relating to problems in telephone network modernization: Control theoretic formulations," INRS-Telecommun., Tech. Rep. 79-07, Oct. 1979.

[8] L. G. Mason, "Network modernization with capital budget constraints," in *Proc. 2nd Int. Symp. Network Planning,* Brighton, England, Mar. 1983.

[9] B. A. Murtaugh and M. A. Saunders, "Large-scale linearly constrained optimization," *Math. Programming,* vol. 14, pp. 41–72, 1978.

★

Lorne G. Mason, for a photograph and biography, see p. 796 of the July 1984 issue of this TRANSACTIONS.

A Class of Reduced-Complexity Viterbi Detectors for Partial Response Continuous Phase Modulation

ARNE SVENSSON, STUDENT MEMBER, IEEE, CARL-ERIK SUNDBERG, SENIOR MEMBER, IEEE, AND TOR AULIN, SENIOR MEMBER, IEEE

Abstract—**Partial response continuous phase modulation (CPM) gives constant envelope digital modulation schemes with excellent power spectra. Both narrow main lobe and low spectral tails can be achieved. When these signals are detected in an optimum coherent maximum likelihood sequence detector (Viterbi detector), power efficient schemes can also be designed, sometimes at the expense of receiver complexity.**

This paper describes a general class of simple Viterbi detectors with reduced complexity compared to the optimum case. The key idea is that the approximate receiver is based on a less complex CPM scheme than the transmitted scheme. The asymptotically optimum reduced-complexity receiver is found for a variety of transmitted schemes and various complexity reduction factors, for a specific class of receivers and modulation indexes. A new distance measure is introduced for the performance analysis. Smooth schemes based on raised cosine pulses are analyzed and simulated for the case of simplified reception. A graceful performance degradation occurs with the reduction of complexity.

I. Introduction

DUE to limited available radio frequency spectrum, spectrally efficient digital modulation schemes have recently gained increased attention. For some applications, notably where nonlinearities are present in repeaters and power amplifiers, it is favorable if the signaling scheme can be of the constant envelope type [8]. It has recently been demonstrated that constant envelope digital modulation methods with excellent power spectra can be constructed; see, for example, [2], [16], [18]. A wide class of such schemes are contained among the continuous phase modulation (CPM) schemes. By using partial response techniques and keeping the information carrying phase continuous, the power spectra become very attractive; see examples in [2], [18]. Partial response digital FM and correlative phase modulation are alternative notations for the CPM schemes.

It has also been shown that schemes with combined good spectral properties and detection properties can be obtained in the CPM family [2]. The schemes with attractive combined spectra and error probability often require a maximum likelihood sequence detector [11] implemented by means of a Viterbi detector. The optimum receiver exploits the structure of the transmitted phase trajectories, and uses the Viterbi algorithm to perform maximum likelihood sequence estimation (MLSE). These receivers typically consist of a filter bank followed by a Viterbi processor, where sometimes the number of states is quite large [2], [7], [9], [17].

In this paper a new class of reduced-complexity Viterbi detectors is considered. It is shown that good performance can be obtained with a Viterbi receiver, where the number of filters and states is reduced. The key idea is to approximate the phase tree (the ensemble of all possible transmitted phase versus time functions) with a phase tree based on a shorter frequency pulse, and use the Viterbi receiver for this shorter scheme instead as a receiver. This scheme, used by the receiver, can be optimized to give the best possible performance.

Throughout the paper, coherent detection and perfect carrier recovery and synchronization are assumed. It is also assumed that the channel is an additive Gaussian noise channel.

II. The Modulation Scheme and the Complexity Reduction Concept

CPM is a digital modulation scheme with constant envelope and continuous phase. The transmitted sequence is an infinite sequence of statistically independent M-ary data symbols with equal probability. All the information is contained in the transmitted phase. The instantaneous frequency is a linear sum of overlapping frequency pulses, where the amplitude of the pulses depends on the data symbols. The length in symbol intervals and the shape of the pulses determines the performance of the modulation system. In this paper

Paper approved by the Editor for Communication Theory of the IEEE Communications Society for publication after presentation in part at the National Telecommunications Conference, New Orleans, LA, November 1981, and at the Global Telecommunications Conference (GLOBECOM), San Diego, CA, November 1983. Manuscript received April 20, 1983; revised February 1, 1984. This work was supported by the Swedish Board of Technical Development under Grant 81-3480.

A. Svensson is with the Institute of Telecommunication Theory, University of Lund, S-220 07 Lund, Sweden.

C.-E. Sundberg was with the Institute of Telecommunication Theory, University of Lund, S-220 07 Lund, Sweden. He is now with AT&T Bell Laboratories, Holmdel, NJ 07733.

T. Aulin is with the Division of Information Theory, Chalmers University of Technology, S-412 96 Gothenburg, Sweden.

0090-6778/84/1000-1079$01.00

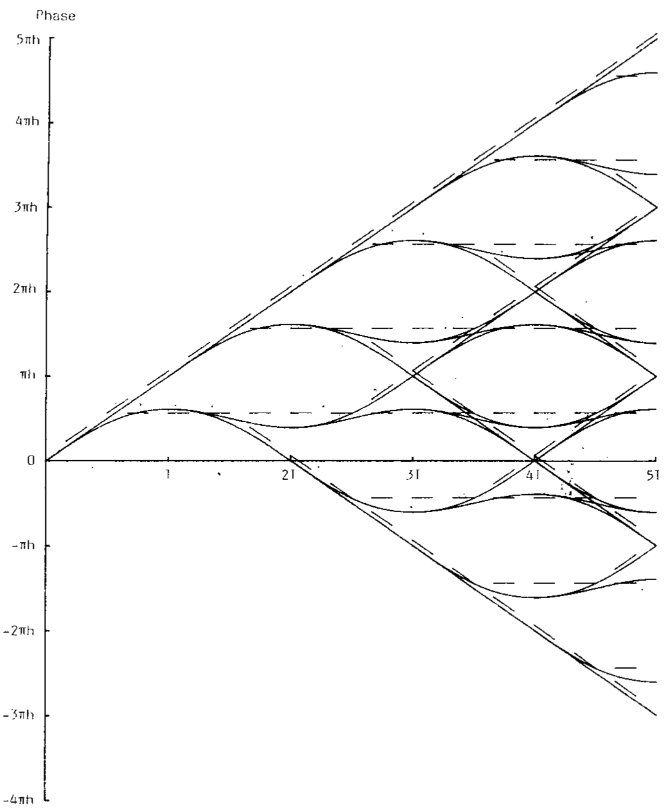

Fig. 1. Transmitter phase tree (solid) for 3RC and receiver phase tree for 2REC. Correct timing, small phase offset.

the frequency pulse is of rectangular (REC) or raised cosine (RC) shape. Since the frequency pulse is time limited, the transmitted phase can be divided into two terms [2], [5], [7]. One term depends only on the last L symbols, where LT is the length of the frequency pulse, and the other term is the sum of the previous symbols. Both terms are multiplied by a constant linearly dependent on the modulation index h. For rational values of h, the second term, taken modulo 2π, only takes a finite number of values. For these h-values the phase tree reduces to a phase trellis. The states in the trellis are defined by the phase state and the correlative state vector, i.e., the prehistory symbols [2], [4], [5], [7]. The phase state definition is, however, not unique. The transmitted phase in a specific symbol interval only depends on the phase state, the correlative state vector, and the last transmitted data symbol. The total number of states in the trellis is $S = p \cdot M^{(L-1)}$ where p is the number of phase states, M is the number of levels, and L is the length of the frequency pulse in symbol intervals. In [2], [7] it is shown that the log likelihood function can be calculated recursively by filtering and can be used as a metric in a Viterbi detector. The receiver consists mainly of a bank of filters and a Viterbi processor. The total number of filters is $F = 2 \cdot M^L$.

The class of coherent suboptimum receivers considered in this paper is also of the maximum likelihood sequence estimating type. The principal idea is the following. The receiver uses a shorter pulse than the transmitter, thus approximating the phase tree. The receiver frequency pulse $g_R(t)$ must be chosen in such a way that the tree generated by $g_R(t)$ is a reasonably good approximation of the phase tree generated by the transmitter frequency pulse $g_T(t)$. The receiver based on $g_R(t)$ is built according to the optimum receiver principles. It is, of course, the optimum receiver for the case of $g_T(t) = g_R(t)$. If the transmitter pulse length in symbol intervals is L_T and the receiver pulse length is L_R, then the complexity reduction factor is $M^{(L_T - L_R)}$ in terms of both the number of receiver states and that of receiver filters.

Fig. 1 shows a transmitter tree based on 3RC and a simplified receiver with a phase tree based on the shorter 2REC pulse. The two trees are shown with a small phase offset relative to the best alignment. Thus, the complete trees can be seen. From Fig. 1 it is clear that the 2REC tree, with properly chosen phase and time offset ($T/2$ in this case), approximates the 3RC scheme fairly well. The quality of this approximation is analyzed in the following sections of this paper. The complexity reduction in this example is a factor of two.

Several interesting observations can be made by studying Fig. 1. For example, in the general case a certain transmitted sequence of binary symbols, e.g., $+1, -1, -1, +1, \cdots$, does not in general correspond to the same phase path in the transmitter and receiver phase trees. This is, of course, a consequence of the approximation made. Thus, in some cases, even if there is no noise at all in the transmission link, there is no perfect match between the transmitted phase path and the correct phase path as generated by the receiver. We will say that the transmitter and receiver signal sets are mismatched. This mismatch should, of course, not be too large, because then detection errors are also made without noise. It is also clear from the example above that the phase synchronization and timing between the transmitter and receiver trees is of great importance. This problem will be addressed below.

III. The Performance Measure

In [1], [2] the minimum normalized squared Euclidean distance is used as a performance measure. For the optimum detector the minimum normalized squared Euclidean distance in the signal space is [1], [2]

$$\frac{D_{\min}^2}{2E_b} \doteq d_{\min}^2 = \min_{\substack{\alpha,\beta \\ \alpha_0 \neq \beta_0}} \left\{ \frac{1}{2E_b} \int_0^{NT} [s(t,\alpha) - s(t,\beta)]^2 \, dt \right\} \tag{1}$$

where $s(t, \alpha)$ is the transmitted CPM signal, E_b is the bit energy, NT is the length of the observation interval, and α and β are data sequences which are different in the first symbol. The distance depends on the difference sequence $\gamma = \alpha - \beta$ rather than the individual data sequences themselves. Assuming $2\pi f_0 T \gg 1$, (1) can be rewritten

$$d_{\min}^2 = \min_{\gamma, \gamma_0 \neq 0} \left\{ \log_2(M) \frac{1}{T} \int_0^{NT} [1 - \cos(\varphi(t, \gamma))] \, dt \right\} \tag{2}$$

where $\varphi(t, \gamma)$ is the transmitted phase, $\gamma_i \doteq \alpha_i - \beta_i$, and M is the number of transmitted levels. Note that the normalization is made in bit energy E_b. It is well known that the error probability behavior for large signal-to-noise ratios (SNR's) is given by the minimum Euclidean distance for coherent detection and a Gaussian channel. The computational effort in computing the minimum distance with a brute-force method is unrealistic even for fairly small values of M and N. One of the key properties of the minimum distance is that an upper bound can easily be calculated [1], [2]. The method of calculating this bound is based on the observation that merges occur in the phase tree for all h-values.

Let us now consider the reduced-complexity receiver. Here a distance measure between mismatched signal sets is introduced [6]. By using this new minimum Euclidean distance, the error probability behavior for large signal-to-noise ratios can be estimated for a given signal and a suboptimum

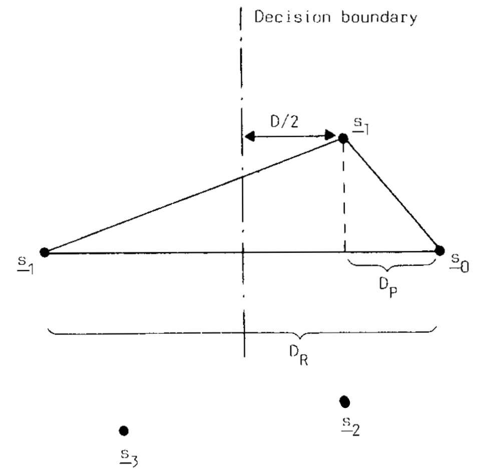

Fig. 2. Calculated distances in the signal space. (Underbar corresponds to boldface in text.)

reduced-complexity receiver. The receiver is assumed to correlate the received signal with all alternatives in the receiver library of signals and choose the alternative with the largest correlation. The receiver principle and the critical parameters are illustrated in Fig. 2.

Let the transmitted signal for a given $\boldsymbol{\alpha}$ be

$$s_T(t, \boldsymbol{\alpha}) = \sqrt{\frac{2E}{T}} \cos\left[\omega_0 t + \varphi(t, \boldsymbol{\alpha})\right] \tag{3}$$

Below we are considering all signals over a time interval of length NT. The receiver generates its library of possible signal alternatives (where $\boldsymbol{\alpha}$ is varying)

$$s_R(t, \boldsymbol{\alpha}) = \sqrt{\frac{2E}{T}} \cos\left(\omega_0 t + \psi(t, \boldsymbol{\alpha})\right). \tag{4}$$

These signals are, of course, also viewed over the same time interval as the transmitted signal. $\varphi(t, \boldsymbol{\alpha})$ and $\psi(t, \boldsymbol{\alpha})$ are calculated from $g_T(t)$ and $g_R(t)$, respectively. $\psi(t, \boldsymbol{\alpha})$ is based on a shorter frequency response than $\varphi(t, \boldsymbol{\alpha})$.

Both the transmitter and the receiver signals are represented in one signal vector space. $\boldsymbol{s}_T$ denotes the transmitted signal point in this context. Let $\boldsymbol{s}_0$ represent the receiver alternative which corresponds to the correct sequence $\boldsymbol{\alpha}$, i.e., the transmitted data sequence. Let $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \cdots$ represent other receiver alternatives. Let the signal alternative $\boldsymbol{s}_1$ correspond to the signal

$$s_R(t, \boldsymbol{\beta}) = \sqrt{\frac{2E}{T}} \cos\left(\omega_0 t + \psi(t, \boldsymbol{\beta})\right). \tag{5}$$

The receiver selects the signal alternative which corresponds to the largest correlation. This corresponds to the signal point which has the smallest distance to the received signal point. Fig. 2 shows the situation in the noiseless case. With additive white Gaussian noise, the signal point $\boldsymbol{s}_T$ is added to a noise vector. When the magnitude of the noise vector is such that the received signal point is located on the other side of the decision boundary, an error occurs. Thus, the distance parameter D is of interest for all pairs of receiver signals $(\boldsymbol{s}_0, \boldsymbol{s}_j)$ $j = 1, \cdots$ for a given transmitted signal $\boldsymbol{s}_T$. Note that D is the distance in the $\boldsymbol{s}_j - \boldsymbol{s}_0$ direction, which is a vector in the receiver signal space. Alternatively, it is possible to project the transmitted signals $\boldsymbol{s}_T$ into the receiver signal space and calculate the distance in this space instead. For large signal-to-noise ratios the smallest D is of particular interest. This is the minimum Euclidean distance for the mismatched signal set. This parameter determines the error probability behavior for large signal-to-noise ratios.

Simple geometry yields that [6]

$$D = D_A{}^2 / D_R \tag{6}$$

where

$$D_R{}^2 = |\boldsymbol{s}_0 - \boldsymbol{s}_1|^2 \tag{7}$$

is the Euclidean distance between the receiver alternatives and $|\cdot|$ is a norm in the Euclidean space; cf. (1). This is referred to as the receiver distance below.

The numerator $D_A{}^2$ can be shown to be [6]

$$D_A{}^2 = |\boldsymbol{s}_1 - \boldsymbol{s}_T|^2 - |\boldsymbol{s}_0 - \boldsymbol{s}_T|^2. \tag{8}$$

This is by definition a positive quantity. In the following the normalized squared distances will be used, i.e.,

$$d^2 = D^2/2E_b = (D_A{}^2/D_R)^2/2E_b = d_A{}^4/d_R{}^2 \tag{9}$$

where $d_R{}^2 = D_R{}^2/2E_b$ and $d_A{}^2 = D_A{}^2/2E_b$.

For comparisons between multilevel systems, the normalization in (9) will be made with the bit energy E_b.

From the analysis above, the receiver performance for large SNR's is determined by the minimum distance $d_{\min}{}^2$ where the minimization is carried out both over transmitter sequences and receiver sequences. Note that for the special case of no mismatch, i.e., $\varphi(t, \boldsymbol{\alpha}) = \psi(t, \boldsymbol{\alpha})$, we have $d_A{}^2 = d_R{}^2$ and, thus, $d^2 = d_R{}^2$. In this case, the error probability is determined by the minimum distance in the signal set, i.e., $d_R{}^2$. Using (1)-(3) in (9) and assuming $\omega_0 T \gg 1$ gives

$$d_A{}^2 = \log_2(M) \cdot \left\{ \frac{1}{T} \int_0^{NT} \cos\left[\varphi(t, \boldsymbol{\alpha}) - \psi(t, \boldsymbol{\alpha})\right] dt - \frac{1}{T} \int_0^{NT} \cos\left[\varphi(t, \boldsymbol{\alpha}) - \psi(t, \boldsymbol{\beta})\right] dt \right\} \tag{10}$$

and

$$d_R{}^2 = \log_2(M) \cdot \frac{1}{T} \int_0^{NT} \left(1 - \cos\left(\psi(t, \boldsymbol{\beta}) - \psi(t, \boldsymbol{\alpha})\right)\right) dt. \tag{11}$$

Equations (9)-(11) are the key to the analysis of the behavior of mismatched signal sets and, thus, reduced-complexity Viterbi receivers for large SNR's. $d_{\min}{}^2$, the minimum of d^2, gives the asymptotic performance of the considered signal with the given receiver. Thus, for large signal-to-noise ratios, the error probability behavior is given by

$$P_e \sim Q\left(\sqrt{\frac{E_b \cdot d_{\min}{}^2}{N_0}}\right) \tag{12}$$

where $Q(\cdot)$ is the error function associated with the normal distribution given by [8]

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy. \tag{13}$$

By studying (10) and (11) we note that $d_A{}^2$ and $d_R{}^2$ can be calculated recursively when N is increased [1], [2], [4]. This can be used for computing d^2 efficiently.

It is clear that the above reduced-complexity receiver is suboptimum. An optimum receiver is matched to the pulse $g_T(t)$, and thus, the squared Euclidean distance is

$$d_T{}^2 = \log_2(M)\left\{N - \frac{1}{T}\int_0^{NT} \cos\left[\varphi(t,\boldsymbol{\alpha}) - \varphi(t,\boldsymbol{\beta})\right]\, dt\right\}. \tag{14}$$

The minimum Euclidean distance for the suboptimum receiver cannot be larger, since in that case the suboptimum receiver performs better than the optimum receiver for large signal-to-noise ratios. Thus, we have the inequality

$$\min(d^2) = \min(d_A{}^4/d_R{}^2) \leqslant \min(d_T{}^2). \tag{15}$$

It is interesting to note that it is possible to have

$$\min(d^2) = \min(d_A{}^4/d_R{}^2) > \min(d_R{}^2) \Leftrightarrow d_A > d_R \quad \text{for that specific } \boldsymbol{\alpha} \text{ and } \boldsymbol{\beta}. \tag{16}$$

This implies that by starting from an optimum scheme with matched receiver and transmitter, it is possible to obtain a larger minimum Euclidean distance by exchanging the transmitter for another one with a larger $d_T{}^2$. The resulting distance is, of course, smaller than $d_T{}^2$ but the reduction depends upon the receiver. The inequality in (15) has been motivated by the discussion above. The inequality seems difficult to prove strictly mathematically, since it is an inequality between two minima.

IV. Properties of the Minimum Euclidean Distance for Mismatched Signal Sets

The properties of the distance measure introduced in Section III will now be investigated more precisely. First we will look into the usefulness of difference sequences [2]. Then upper bounds $d_B{}^2$ on the minimum distance are calculated.

For calculation of Euclidean distances for optimum receivers, only the phase difference $\varphi(t,\boldsymbol{\alpha}) - \varphi(t,\boldsymbol{\beta}) = \varphi(t,\boldsymbol{\alpha}-\boldsymbol{\beta})$ has to be considered. This simplifies the calculations since only the difference sequence $\boldsymbol{\gamma} = \boldsymbol{\alpha} - \boldsymbol{\beta}$ has to be considered. It will now be investigated if and when difference sequences also can be used for calculation of Euclidean distances for mismatched receivers. The crucial quantities which have to be investigated are $\varphi(t,\boldsymbol{\alpha}) - \psi(t,\boldsymbol{\alpha})$ and $\varphi(t,\boldsymbol{\alpha}) - \psi(t,\boldsymbol{\beta})$; see (10). The first difference is

$$\begin{aligned}\varphi(t,\boldsymbol{\alpha}) - \psi(t,\boldsymbol{\alpha}) &\triangleq \epsilon(t,\boldsymbol{\alpha}) \\ &= 2\pi h \sum_i \alpha_i\left[q_T(t-iT) - q_R(t-iT)\right] \\ &= 2\pi h \sum_i \alpha_i q_\Delta(t-iT)\end{aligned} \tag{17}$$

where

$$q_\Delta(t) = q_T(t) - q_R(t) = \int_{-\infty}^{t} \left[g_T(\tau) - g_R(\tau)\right]\, d\tau. \tag{18}$$

The symbol timing between the transmitter and the receiver is defined by letting

1) $g_R(t)$ occupy the interval $[0, L_R \cdot T]$, and
2) $g_T(t)$ occupy the interval $[-L_\Delta \cdot T, (L_T - L_\Delta)\cdot T]$

where $L_T \geqslant L_R$ and

$$L_\Delta = \frac{L_T - L_R}{2}. \tag{19}$$

By proper phase synchronization ψ_0, this will assure the best fit between the two phase trees given by the ensemble of $\varphi(t,\boldsymbol{\alpha})$ and $\psi(t,\boldsymbol{\alpha})$, respectively. The properties of $q_R(t)$ and $q_T(t)$ lead to $q_\Delta(t) \equiv 0$ when $t \leqslant -L_\Delta T$ and $t \geqslant (L_T - L_\Delta)T$. Thus, it can be seen that the dependence of $\boldsymbol{\alpha}$ in $\epsilon(t,\boldsymbol{\alpha})$ in the interval $nT \leqslant t \leqslant (n+1)T$ lies in the $(2\Lambda_\Delta + L_R)$-ary vector

$$(\alpha_{n-L_R+1-\Lambda_\Delta}, \cdots, \alpha_n, \cdots, \alpha_{n+\Lambda_\Delta}) = \boldsymbol{\alpha}_{n+\Lambda_\Delta} \tag{20}$$

where

$$\Lambda_\Delta = \left\lceil \frac{L_T - L_R}{2} \right\rceil = \text{smallest integer larger than or equal to } \frac{L_T - L_R}{2}. \tag{21}$$

Now the expression $\varphi(t,\boldsymbol{\alpha}) - \psi(t,\boldsymbol{\beta})$ is considered. By constraining t to the interval $nT \leqslant t < (n+1)T$ and using the properties of $q_T(t)$ and $q_R(t)$, we have

$$\begin{aligned}\varphi(t,\boldsymbol{\alpha}) - \psi(t,\boldsymbol{\beta}) &= h\pi \sum_{i \leqslant n-L_R-\Lambda_\Delta} (\alpha_i - \beta_i) - h\pi \sum_{i=n-L_R-\Lambda_\Delta+1}^{n-L_R} \beta_i \\ &\quad + 2\pi h \sum_{i=n-L_R-\Lambda_\Delta+1}^{n-L_R} \alpha_i q_T(t-iT) \\ &\quad + 2\pi h \sum_{i=n-L_R+1}^{n} \left[\alpha_i q_T(t-iT) - \beta_i q_R(t-iT)\right] \\ &\quad + 2\pi h \sum_{i=n+1}^{n+\Lambda_\Delta} \alpha_i q_T(t-iT)\end{aligned} \tag{22}$$

and only the first term depends on $\boldsymbol{\alpha} - \boldsymbol{\beta}$. For the other terms the two $(2\Lambda_\Delta + L_R)$-ary vectors $\boldsymbol{\alpha}_{n+\Lambda_\Delta}$ and $\boldsymbol{\beta}_{n+\Lambda_\Delta}$, defined according to (20), must be used.

Thus, for an observation interval $N \leqslant L_R + \Lambda_\Delta$ symbol intervals, both the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ sequences must be used to obtain all the Euclidean distances. For $N > L_R + \Lambda_\Delta$ symbol intervals, however, the phase separation created by the initial $N - L_R - \Lambda_\Delta$ data symbols can be expressed by means of the sum of the difference sequence $\boldsymbol{\alpha} - \boldsymbol{\beta}$.

Upper Bound $d_B{}^2(h)$

The minimum Euclidean distance for optimum receivers has been shown to be upper bounded [1], [2]. This holds for all observation lengths N, including infinity. The upper bound is calculated by identifying so-called merges in the phase tree.

For mismatched receivers, merges do not exist in general, at least not merges of the kind defined in [2]. It is possible, however, to choose sequences $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such that the calculation of $d_A{}^2$ only has to be performed over a finite interval, although $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are infinitely long. The sequences $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are chosen exactly as the merge sequences given in [2], [7], e.g.,

$$\begin{cases} \boldsymbol{\alpha} = \cdots \alpha_{-2}, \alpha_{-1}, +1, -1, \alpha_2, \cdots \\ \boldsymbol{\beta} = \cdots \beta_{-2}, \beta_{-1}, -1, +1, \beta_2, \cdots \end{cases} \tag{23}$$

where $\alpha_i = \beta_i$ when $i \neq 0$, 1 for the first merge in the binary case. Now this pair of sequences yields a merge at $t = (L_R + 1)T$ in the phase tree given by the ensemble of $\psi(t, \boldsymbol{\alpha})$. Now the expression for $d_A{}^2$ in (10) will be studied for this pair of sequences and an infinite observation length $-\infty < t < \infty$. Since $\alpha_i = \beta_i$, $i < 0$, it is clear that $\psi(t, \boldsymbol{\alpha}) = \psi(t, \boldsymbol{\beta})$ for all $t \leqslant 0$. For all $t > (L_R + 1)T$, we know from the previous discussion about difference sequences that $\psi(t, \boldsymbol{\alpha}) = \psi(t, \boldsymbol{\beta})$ since the influence of α_0, α_1 and β_0, β_1 does not affect the shape of these two phase trajectories. The separation between these two phase trajectories is affected by these symbols. This separation equals zero since $\Sigma_i \alpha_i = \Sigma_i \beta_i$, and thus, $\psi(t, \boldsymbol{\alpha}) = \psi(t, \boldsymbol{\beta})$ for all $t \leqslant 0$ and $t \geqslant (L_R + 1)T$. This is equivalent to

$$\varphi(t, \boldsymbol{\alpha}) - \psi(t, \boldsymbol{\alpha}) = \varphi(t, \boldsymbol{\alpha}) - \psi(t, \boldsymbol{\beta}) \qquad t \leqslant 0, \quad t \geqslant (L_R + 1)T \tag{24}$$

and therefore, the two integrals in (10) are equal over these intervals. Thus, it is sufficient to calculate $d_A{}^2$ over the time interval where the trajectories in the receiver phase tree are unmerged.

In the M-ary case there are more merge sequences, and of course, there are also merges later than the first. In some cases the upper bound on d^2 can be improved by also taking such merges into account. Unlike the case for matched receivers, the upper bound can also depend upon a finite number of the prehistory data symbols $\alpha_{-1}, \alpha_{-2}, \cdots$ and $\beta_{-1}, \beta_{-2}, \cdots$. Only one prehistory sequence must be considered, however, since $\alpha_i = \beta_i$; $i < 0$, and not a pair of sequences.

V. The Optimized Reduced-Complexity Receiver

The aim of this section is to find the optimum reduced-complexity receiver. In this case a natural function to optimize is the error probability at a given signal-to-noise ratio. From the previous sections it is known that this is unrealistic. Therefore, the minimum Euclidean distance has been chosen instead. If the minimum Euclidean distance is optimized, the asymptotic error probability is also optimized. Thus, the receiver frequency pulse $g_R(t)$ that maximizes the minimum Euclidean distance has to be found. The function which is to be maximized is therefore given by (9)-(11). An analytic solution to this seems difficult to find. Therefore, a numerical maximization is done instead. With a numerical maximization, it would in principle be possible to find the optimum reduced-complexity receiver, to a given transmitted scheme, a given modulation index, and a given complexity reduction, for every observation length of N symbol intervals. The computational efforts to do this are, however, very large. Therefore, the upper bound on the minimum Euclidean distance is maximized instead. A tight upper bound can always be found, and therefore, it is possible to find an observation length sufficiently large for the minimum distance to reach this upper bound.

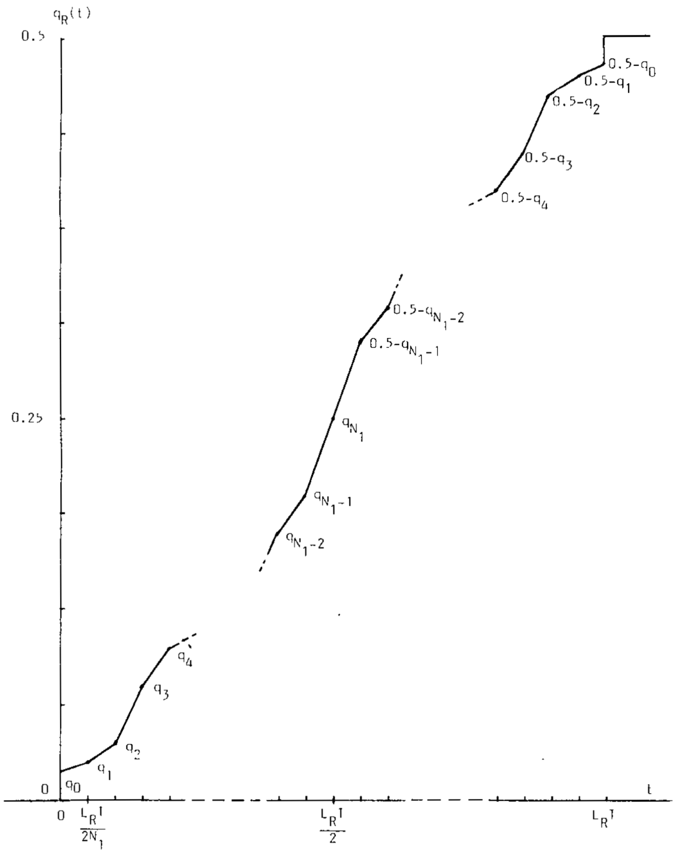

Fig. 3. The piecewise linear phase response of the reduced-complexity receiver.

In this paper we have chosen the phase response $q_R(t)$ to be a piecewise linear function for $0 < t < L_R T$ which is allowed to be discontinuous at $t = 0$ and at $t = L_R T$.

$$q_R(t) = \begin{cases} 0; & t \leqslant 0 \\ q_i + \dfrac{2N_1(q_{i+1} - q_i)}{L_R T} \cdot \left(t - i\dfrac{L_R T}{2N_1}\right); & i \cdot \dfrac{L_R T}{2N_1} < t \leqslant (i+1)\dfrac{L_R T}{2N_1}, \\ & i = 0, 1, \cdots, N_1 - 1 \\ \dfrac{1}{2} - q_R(L_R T - t); & \dfrac{L_R T}{2} < t < L_R T \\ \dfrac{1}{2}; & t \geqslant L_R T. \end{cases} \tag{25}$$

$q_{N_1} = 1/4$ by definition; see also Fig. 3.

Since the transmitter pulse $g_T(t)$ is symmetric around $t = L_T \cdot T/2$, it is assumed that the receiver pulse $g_R(t)$ is also symmetric around $t = L_R \cdot T/2$. Maximizations without this assumption have also shown that this is most probably the case. This means that the interval $0 \leqslant t \leqslant L_R T/2$ is divided into N_1 subintervals, over which $q_R(t)$ is linear. The variables q_i are the values at the endpoints of these intervals and $q_{N_1} = q_R(L_R T/2) = 1/4$ due to symmetry. Then $q_R(t)$ for $L_R T/2 < t \leqslant L_R T$ is given by symmetry. Now the set of $q_i, i = 0, \cdots, N_1 - 1$ maximizing the upper bound has to be found. This is done by a steepest descent algorithm [5], [19]. The results show that in general $q_0 \neq 0$ yields the optimum. By choosing N_1 large, it is possible to approximate every continuous phase response $q_R(t)$ (except at $t = 0$ and $t = L_R T$) that corresponds to a symmetric frequency pulse $g_R(t)$.

VI. Numerical Results

In this section, both distance results and simulated error probability results will be given.

Distance Results

The distance measure introduced in Section III is used. The observation interval length is always measured in the receiver tree. The results are presented as the minimum normalized squared Euclidean distance as a function of modulation index h and observation interval length N. This is done for a number of pairs of transmitter and receiver pulse shapes $g_T(t)$ and $g_R(t)$. Results are given both for optimum pulses $g_R(t)$ and for ad hoc chosen pulses. Comparisons will always be made to the optimum receiver based on $g_T(t)$. In [3]-[5] a variety of distance results are given for different receivers.

Considering the results for the optimum reduced-complexity receivers, note that the scheme is optimum only for a specific modulation index. In these graphs the upper bound for the optimum receiver, the upper bound for the optimum reduced-complexity receiver, and the upper bound for an ad hoc reduced-complexity receiver are also given. The optimum reduced-complexity receiver schemes will be denoted L_TRC – $L_R Th$, where L_T and L_R are the length in symbol intervals of the transmitter and receiver pulse, respectively, and h is the modulation index at which this receiver is optimum. L_TRC is the transmitted scheme to which this receiver pulse is optimum. The ad hoc reduced-complexity receivers will in the same manner be referred to as transmitted scheme-receiver scheme, e.g., 3RC-1REC. For all the piecewise linear pulses considered, the maximization is done with $N_1 = 10$, i.e., the interval $0 \leqslant t \leqslant L_R T$ is divided into 20 subintervals.

Fig. 4 shows a scheme with a complexity reduction of 2, i.e., 3RC-2T0.75. For comparison, the upper bound for the 3RC-3RC2 scheme [5] and the upper bound for the optimum 3RC receiver are given. From [5] it is known that the phase tree given by the optimum $g_R(t)$ pulse approximates the 3RC phase tree very well. In the distance graph it is also seen that the distance is very close to the upper bound for the optimum 3RC receiver. The loss at $h = 3/4$ compared to optimum 3RC is small for both reduced-complexity receivers. Note that this receiver phase response for 3RC-2T0.75 is not optimum for larger h-values, as seen in Fig. 4.

The degradation compared to the optimum scheme for the reduced-complexity receivers optimum at other modulation indexes is very close to the degradation for 3RC-2T0.75 [5].

Now, transmitted 4RC is considered instead. Fig. 6 gives the results for 4RC-2T0.50. With the complexity reduction factor of 4 in mind, the phase tree, generating the receiver signal set of the optimum reduced-complexity receiver, approximates the 4RC phase tree fairly well. This phase tree is shown in Fig. 5, where the 4RC phase tree is also given (dashed). The loss at $h = 1/2$ compared to the optimum 4RC receiver is about 0.35 dB. As seen from the distance graph, this upper bound is reached with an observation length of four symbol intervals.

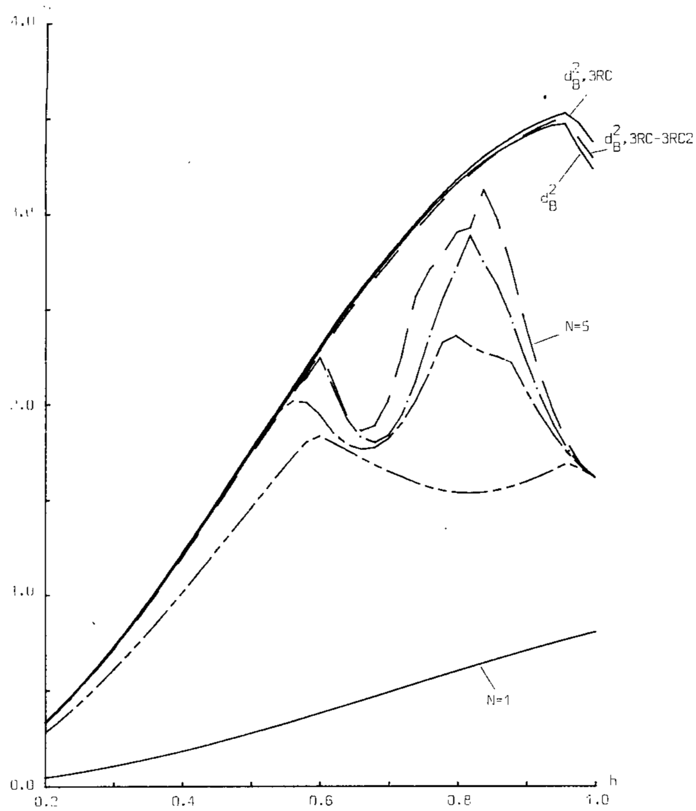

Fig. 4. Minimum distance for binary 3RC-2T0.75. Note that this receiver is optimum only for $h = 3/4$.

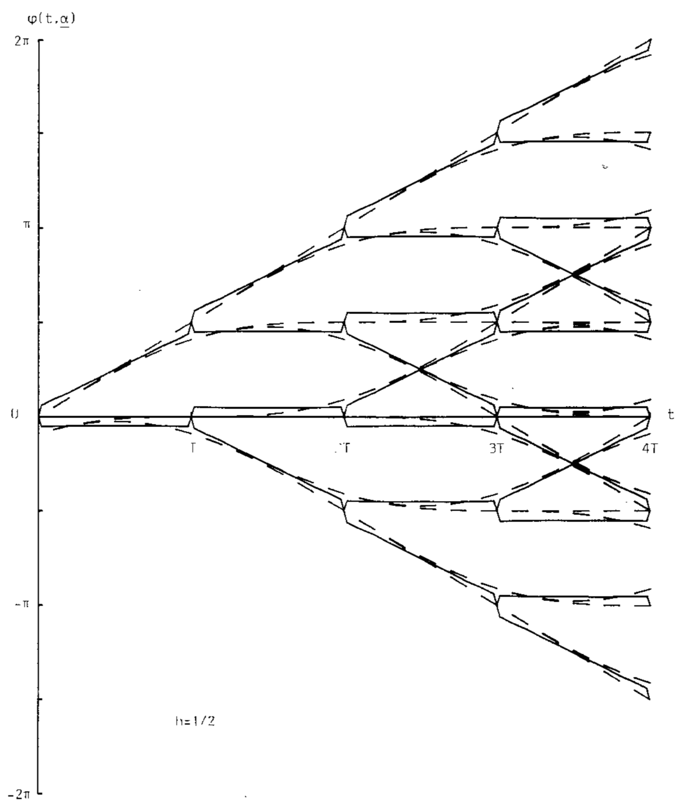

Fig. 5. Phase tree for binary 4RC-2T0.50 (solid) and 4RC (dashed).

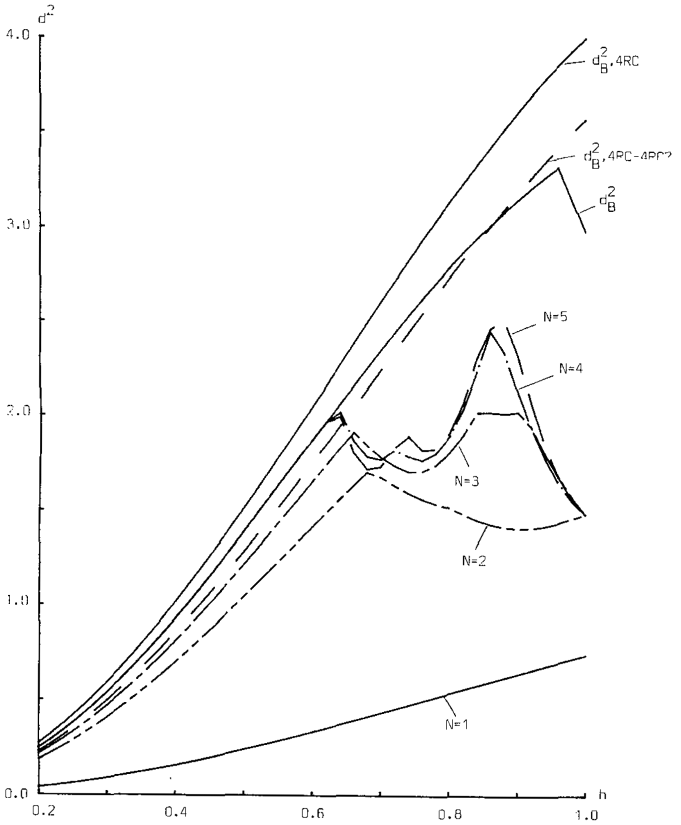

Fig. 6 Minimum distance for binary 4RC-2*T*0.50. Note that this receiver is optimum only for $h = 1/2$.

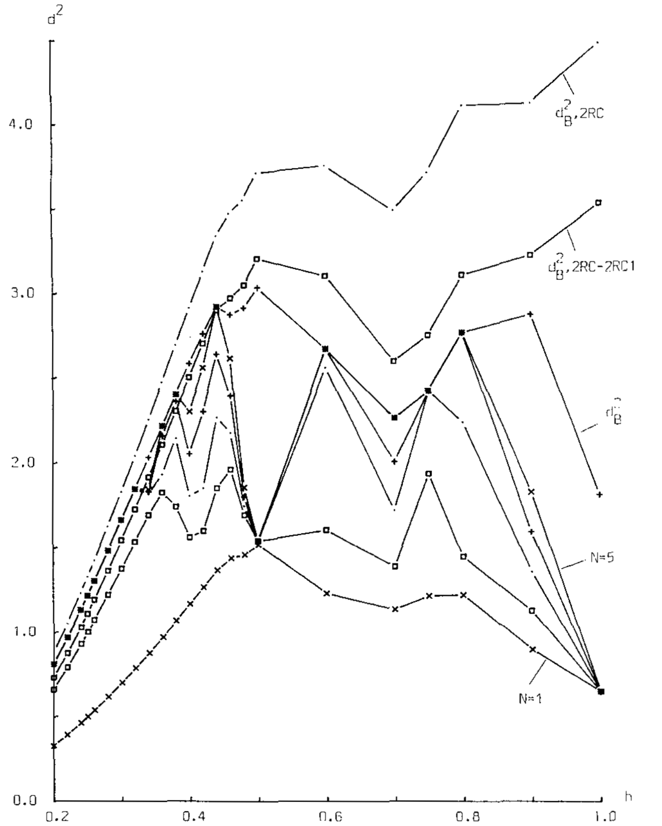

Fig. 7. Minimum distance for quaternary 2RC-1*T*0.25. Note that this receiver is optimum only for $h = 1/4$.

Optimizations and distance calculations for some quaternary ($M = 4$) schemes have also been done. Fig. 7 shows the results for quaternary 2RC-1T0.25. The complexity reduction factor is now 4. Note that the calculated values are marked with •, X, +, or □ and the lines only connect the corresponding values. The distances shown are normalized with the bit energy E_b and are therefore comparable to the binary results. The loss at $h = 1/4$ compared to the optimum 2RC receiver is about 0.35 dB. The upper bound at $h = 1/4$ is reached with an observation length $N = 3$.

Simulated Error Probabilities

The minimum distance calculations give the error probability behavior for large signal-to-noise ratios. For low signal-to-noise ratios it is necessary to perform computer simulations. This is so for the optimum Viterbi receiver also. However, it seems even more appropriate for the optimum reduced-complexity receiver. From the distance considerations in the previous sections it follows that, relatively speaking, the error event corresponding to the minimum Euclidean distance is a more unusual event for the reduced-complexity receiver than for the optimum receiver. A specific prehistory must have occurred in the phase tree for the minimum distance error event to take place. This is not so for the optimum receiver. A minimum distance can occur for any prehistory, since the phase difference is the same independent of the prehistory.

Simulations have been made for transmitted 4RC with $h = 1/2$ and transmitted quaternary 2RC with $h = 1/4$. The receivers used are the optimum Viterbi receiver, the optimum reduced-complexity receiver, and the ad hoc reduced-complexity receiver with the largest $d_{\min}^2$. All the simulation results are based on 2500 errors and the path memory in the Viterbi detector is 20 symbol intervals, which for all the schemes is enough to reach the asymptotic performance given by $d_{\min}^2$. The lower bound on the error probability for the optimum Viterbi receiver [2] is also given for comparison. Note that the simulated error probabilities are marked with X, +, or □ and then connected with a straight line. This is also the case for the lower bound. The SNR in Figs. 8 and 9 is E_b/N_0.

Fig. 8 shows the results for 4RC-4RC2 (+), 4RC-2T0.50 (X), and optimum 4RC (□), together with the optimum lower bound for $h = 1/2$. It is seen that the reduced-complexity receivers perform approximately equally for low SNR's, while the optimum receiver performs about 0.1 dB better at $P = 10^{-2}$. The loss at low SNR's for the reduced-complexity receivers is, however, not at all as large as the minimum distance tells us. This is due to the fact that the probability for the error event given minimum distance to occur is very low.

Finally, Fig. 9 gives the results for some quaternary schemes. These are 2RC-2RC1 (+), 2RC-1T0.25 (X), and optimum 2RC (□) for $h = 1/4$. The lower bound for the optimum scheme is also shown. In this case 2RC-1T0.25 performs in between the other two schemes. Still, the loss compared to the optimum scheme is low. For further numerical results, see [4], [5].

VII. Tradeoff Between Complexity Reduction and Performance Degradation

To illustrate the tradeoff between complexity reduction and performance degradation, we have selected data for some of the schemes in the distance calculation above and in [3]-[5]. By complexity reduction we have defined the reduction factor $M^{(L_T - L_R)}$. The number of states and filters in the suboptimum receiver is reduced by this factor compared to the optimum receiver. By performance degradation we define the factor $10 \circ \log_{10}(d_T^2/d_{\min}^2)$, i.e., the difference in power in decibels between the optimum scheme and the suboptimum

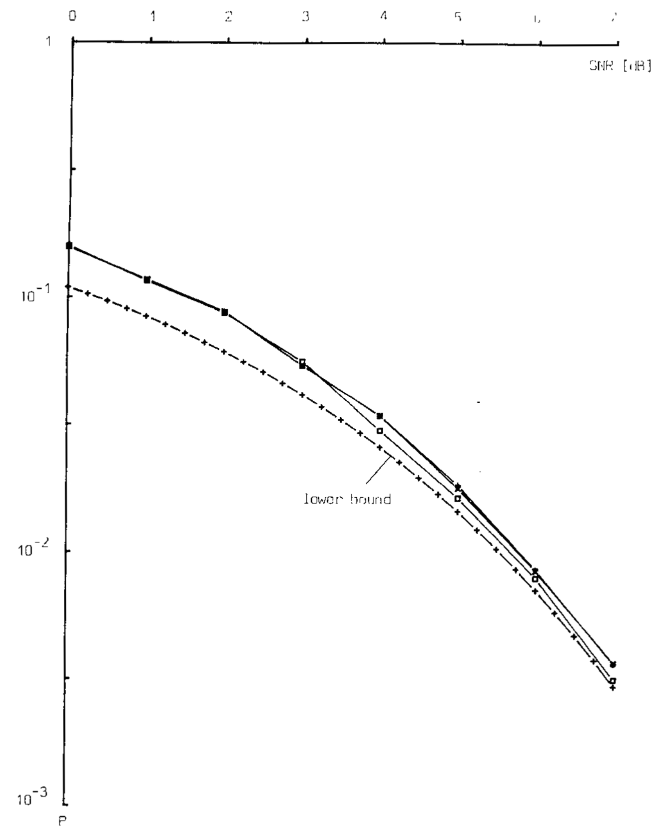

Fig. 8. Simulated bit error probability for 4RC-4RC2 (+), 4RC-2T0.50 (×), and optimum 4RC (□) for h = 1/2. The lower bound for the optimum receiver is also shown. The results are based on 2500 errors.

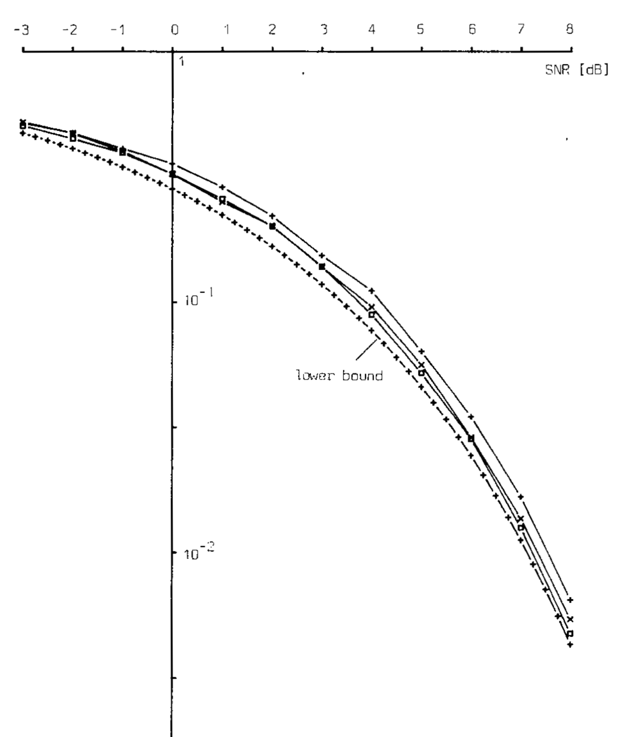

Fig. 9. Simulated symbol error probability for quarternary 2RC-2RC1 (+), 2RC-1T0.25 (×), and optimum 2RC(□) for h = 1/4. The lower bound for the optimum receiver is also shown. The results are based on 2500 errors.

TABLE I
POWER DEGRADATION VERSUS COMPLEXITY REDUCTION FOR THE OPTIMUM REDUCED-COMPLEXITY RECEIVER COMPARED TO THE OPTIMUM RECEIVER FOR h = 3/4

Transmitted scheme	M=2 2RC	M=2 3RC	M=2 4RC	M=2 3RC	M=2 4RC	M=4 2RC	M=2 4RC
L_R	1	2	3	1	2	1	1
Degradation [dB]	0.13	0.02	0.02	0.86	0.36	1.28	2.40
$M^{L_T-L_R}$	2	2	2	4	4	4	8

scheme for large SNR's. This is the same loss or degradation that has been referred to above. The performance degradation varies with h, while the complexity reduction factor only is a function of M and $L_T - L_R$.

Table I shows the asymptotic performance degradation for the optimum reduced-complexity receiver compared to the optimum receiver versus the corresponding complexity reduction factor. The table gives the results for $h = 3/4$. It is seen that the degradation is very small for a low complexity reduction factor, while it is increasing for larger factors.

VIII. DISCUSSION AND CONCLUSIONS

A general method of reducing the complexity of the general optimum receiver structure for partial response CPM (digital FM) has been presented. A new distance measure has been introduced for analyzing the asymptotic performance of schemes with suboptimum receivers. Coherent transmission on the additive white Gaussian noise channel is considered and the distance results give the error probability behavior for large signal-to-noise ratios. For low signal-to-noise ratios, computer simulations have been performed for some selected schemes.

The optimum reduced-complexity Viterbi receiver for partial response continuous phase modulation (digital FM) has also been found. The optimization is performed by means of a numerical maximization of the upper bound on the squared minimum Euclidean distance. The phase tree of the CPM scheme is approximated in the receiver by a phase tree based on a shorter frequency pulse. This shorter pulse is assumed to have a piecewise linear phase response. The piecewise linear phase response, giving a maximum upper bound on the Euclidean distance, is found. The optimum receiver phase response varies with the modulation index h for a given receiver frequency pulse length and a given transmitter frequency pulse. This does not give an optimum reduced-complexity receiver in a global sense, since a subclass of receiver phase responses is allowed in the optimization.

The results show that the number of filters and states can be reduced significantly at the expense of a power degradation of the order of 0.5–1.0 dB. A graceful degradation occurs. The larger the complexity reduction, the larger the performance degradation. Partial response schemes with very smooth power spectra can be received in Viterbi receivers with a rather low number of states.

The reduced-complexity receiver is a robust receiver and the idea of approximating the phase tree can also be applied to other types of smooth frequency responses, e.g., TFM GMSK, SRC [2], [16], [21]. Similar results are expected for these schemes.

It should be noted that the class of schemes analyzed above, in terms of the distance between mismatched signal sets, has the following feature. The power spectrum of the received signal is narrower than the power spectrum of the "optimum" received signal for which the reduced-complexity receiver is designed. A noise-reducing bandpass filter might improve the performance beyond what is indicated by the distance value, which is calculated by means of matched filters in white noise arguments. Simulations should be performed to validate this hypothesis.

Even without a noise-reducing bandpass filter, the distance values for the suboptimum receivers seem to be a little pessimistic for low and intermediate error probability values (signal-to-noise ratios). This could be explained by the rare occurrence of error events (relatively speaking) corresponding to the minimum distance.

An analytic optimization still has not been done and seems very difficult to do. However, these results seem to be very close to an absolute optimum reduced-complexity receiver. A comparison with the results for the ad hoc pulses shows that the minimum distance is not especially sensitive to small variations in the pulse shape.

In [20] a reduced-complexity method is given where the Viterbi decoder has a predetermined processing order and a reduced number of survivor signals. This method is claimed to be asymptotically optimum for certain modulation indexes. So far it has only been applied to very simple piecewise linear phase responses for binary schemes. It is not clear how this method compares to ours for the more interesting cases, i.e., long smoothing pulses for the binary schemes, and to the quaternary $M = 4$ schemes. Further work is required. Another type of reduced-complexity receiver is the very simple MSK-type receiver [21], which is limited to binary schemes with modulation index $h = 1/2$.

With the receiver in this paper the signal spectrum is unaffected, and only receiver complexity is reduced. Thus, the very low sidelobes of, e.g., 3RC can be maintained with simpler detectors. It should be observed that the technique above is quite general and applies to nonbinary CPM schemes and modulation indexes not necessarily 1/2. It also applies to a wide range of smooth pulse shapes, e.g., TFM [16], spectral raised cosine (SRC) [2], CORPSK [12], GMSK [21], etc.

REFERENCES

[1] T. Aulin and C.-E. Sundberg, "Continuous phase modulation—Part I: Full response signaling," *IEEE Trans. Commun.*, vol. COM-29, pp. 196–209, Mar. 1981.

[2] T. Aulin, N. Rydbeck, and C.-E. Sundberg, "Continuous phase modulation—Part II: Partial response signaling," *IEEE Trans. Commun.*, vol. COM-29, pp. 210–225, Mar. 1981.

[3] T. Aulin, C.-E. Sundberg, and A. Svensson, "Viterbi detectors with reduced complexity for partial response continuous phase modulation," in *Nat. Telecommun. Conf. Rec.*, New Orleans, LA, 1981, pp. A7.6.1–A7.6.7.

[4] A. Svensson and C.-E. Sundberg, "Optimized reduced-complexity Viterbi detectors for CPM," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, San Diego,CA, 1983, pp. 22.1.1–22.1.8.

[5] ——, "On optimized reduced-complexity Viterbi detectors for partial response continuous phase modulation," Telecommun. Theory, Univ. of Lund, Lund, Sweden, Tech. Rep TR-174, Nov. 1982.

[6] A. Svensson, C.-E. Sundberg, and T. Aulin, "A distance measure for simplified receivers," *Electron. Lett.*, vol. 19, pp. 953–954, Nov. 10, 1983.

[7] T. Aulin, N. Rydbeck, and C.-E. Sundberg, "Transmitter and receiver structures for *M*-ary partial response FM," in *Proc. Int. Zürich Sem. Digital Commun.*, Zürich, Switzerland, 1980, pp. A.2.1–A.2.6.

[8] N. Rydbeck and C.-E. Sundberg, "Recent results on spectrally efficient constant envelope digital modulation methods," in *Conf. Rec., Int. Conf. Commun.*, Boston, MA, 1979, pp. 42.1.1–42.1.6.

[9] T. Aulin, "Viterbi detection of continuous phase modulated signals," in *Nat. Telecommun. Conf. Rec.*, Houston, TX, 1980, pp. 14.2.1–14.2.7.

[10] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

[11] W. P. Osborne and M. B. Luntz, "Coherent and noncoherent detection of CPFSK," *IEEE Trans. Commun.*, vol. COM-22, pp. 1023–1036, Aug. 1974.

[12] D. Muilwijk, "Correlative phase shift keying—A bandwidth and power efficient envelope modulation," in *Conf. Rec., Int. Conf. Commun.*, 1980, pp. 26.6.1–26.6.8.

[13] G. J. Garrison, "A power spectral density analysis for digital FM," *IEEE Trans. Commun.*, vol. COM-23, pp. 1228–1243, Nov. 1979.

[14] A. Lender, "Correlative level coding for binary data transmission," *IEEE Spectrum*, vol. 3, pp. 104–115, Feb. 1966.

[15] P. Kabal and S. Pasupathy, "Partial response signaling," *IEEE Trans. Commun.*, vol. COM-23, pp. 921–934, Sept. 1975.

[16] F. de Jager and C. B. Dekker, "Tamed frequency modulation, a novel method to achieve spectrum economy in digital transmission," *IEEE Trans. Commun.*, vol. COM-26, pp. 534–542, May 1978.

[17] J. B. Anderson and C. P. Taylor, "A bandwidth efficient class of signal space codes," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 703–712, Nov. 1978.

[18] T. Aulin and C.-E. Sundberg, "An easy way to calculate power spectrum for digital FM," *IEE Proc. Part F*, vol. 130, pp. 519–526, Oct. 1983.

[19] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.*, vol. 6, pp. no. 2, pp. 163–168, 1963.

[20] S. J. Simmons and P. H. Wittke, "Low complexity decoders for constant envelope digital modulations," in *Global Telecommun. Conf. Rec.*, Miami, FL, Nov. 1982, pp. E7.7.1–E7.7.5.

[21] A. Svensson and C.-E. Sundberg, "Optimum MSK-type receivers for CPM on Gaussian and Rayleigh fading channels," *IEE Proc., Part F*, vol. 131, pp. 480–490, Aug. 1984.

★

Arne Svensson (S'82) was born in Vedåkra, Sweden, on October 22, 1955. He received the M.S. degree in electrical engineering from the University of Lund, Lund, Sweden, in 1979, and the Dr.Ing. and Dr. Techn. degrees from the Institute of Telecommunication Theory, University of Lund, in 1982 and 1984, respectively.

His research interests are digital modulation methods with applications to land mobile radio systems. Since September 1982 he has been working as a Research Assistant in this field on a project supported by the Swedish Board of Technical Development.

★

Carl-Erik Sundberg (S'69–M'75–SM'81) was born in Karlskrona, Sweden, on July 7, 1943. He received the M.S.E.E. and Dr. Techn. degrees from the Lund Institute of Technology, University of Lund, Sweden, in 1966 and 1975, respectively.

Currently he is a member of the Technical Staff at AT&T Bell Laboratories, Holmdel, NJ. From 1977 to 1984 he was a Research Professor (Docent) in the Department of Telecommunication Theory, University of Lund. During 1976, he was with the European Space Research and Technology Centre (ESTEC), Noordwijk, The Netherlands, as an ESA Research Fellow. He has held positions as Consulting Scientist at LM Ericsson, SAAB-SCANIA, Sweden, and at Bell Laboratories, Holmdel. His consulting company, SUNCOM, has been involved in studies of error control methods and modulation techniques. His research interests include source coding, channel coding (especially decoding techniques), digital modulation methods, fault-tolerant systems, digital mobile radio systems, spread-spectrum systems, and digital satellite communication systems.

Dr. Sundberg is a member of Svenska Elektroingenjörers Riksförening.

★

Tor Aulin (S'77–M'80–SM'83) was born in Malmö, Sweden, on September 12, 1948. He received the M.S. degree in electrical engineering from the University of Lund, Lund, Sweden, in 1974, and the Dr.Techn. degree from the Institute of Telecommunication Theory, University of Lund, in November 1979.

Since 1979 he has been a Postdoctoral Fellow at the Institute of Telecommunication Theory, during which time he also was a Visiting Scientist at the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute, Troy, NY. After this he spent one year at the European Space Agency (ESA), the European Space Research and Technology Center (ESTEC), Noordwijk, The Netherlands, as an ESA Research Fellow. Currently he is a Research Professor in the Department of Information Theory, with Data Communications, at the Chalmers University of Technology, Gothenburg, Sweden. Among his research interests are digital radio channel characterization, digital satellite communication systems, digital modulation techniques, digital mobile radio, and signal space coding. His company, AUCOM, has performed several advanced studies as a consultant to some of the major international organizations dealing with developing and operating satellite communication systems.